BioFaceNet Report

Doruk Karinca

This is a summary of the text and figures that are already in the Jupyter notebook. There are many implementation-specific notes, assumptions, and observations in the Jupyter notebook's text blocks that are omitted here.

BioFaceNet is a UNet coupled with model-based assumptions about camera and light that decomposes a face into parameter maps, primarily predicting melanin and haemoglobin content at any pixel in the face.

Dataset observations

In Sample #15 in the figure below, the shading plot has wider lips than the original color image where the mouth is slightly open, probably due to overfitting the CelebA dataset where celebrities are predominantly smiling and that's what the network learned in the *Neural Face Editing with Intrinsic Image Disentangling* paper. This case is common whenever the subject isn't smiling or has an open mouth. Sometimes the nose size is rendered differently. Such additional samples are demonstrated in the notebook for reference. This may hint at a dataset bias that potentially affects convergence.

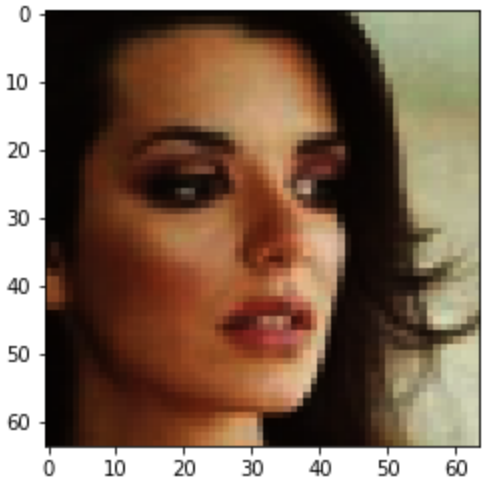

Figure 1a: Original image of sample #15 from CelebA.

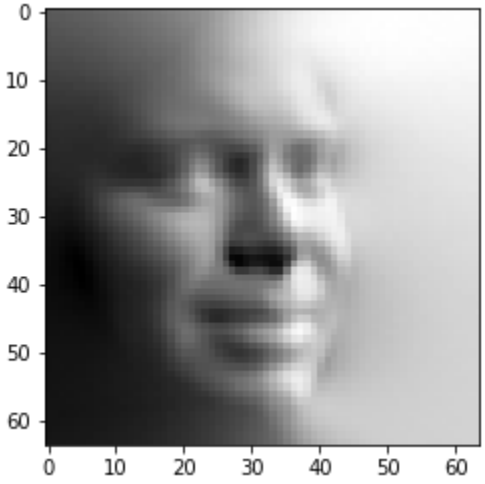

Figure 1b: Shading of sample #15 from CelebA; notice how the mouth shape is not the same as Figure 1a.

Accuracy of BioFaceNet implementation and results

After many iterations overfitting a small dataset, the model could only output skin tone-like noise as seen in the figure below. Trying different weights for the losses, and trying different assumptions during the implementation process did not help. Since these assumptions are closely related to the code, please refer to the Jupyter notebook where text blocks accompanying the code explain these details.

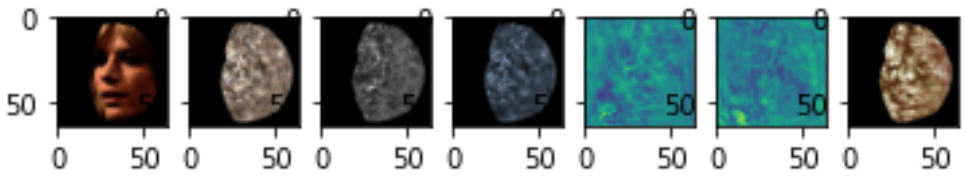

Figure 2a: Face decomposition maps that resemble BioFaceNet's Figure 1. From left to right the plotted maps are original RGB, reconstructed, RGB, shading, specularities, melanin, haemoglobin, and diffuse albedo (as in Figure 1 of the original BioFaceNet paper).

To diagnose the problem, I replace the shading parameter with a grayscale version of the original image i.e. fixing the shading so that the network can easily copy this grayscale, add coloring, and reconstruct the original RGB as we are intentionally overfitting a small dataset below. This is done by passing `copy_shading=True` to UNet. You can visualize the grayscale image in the third column of each sample. Unfortunately this trick still does not help the reconstruction.

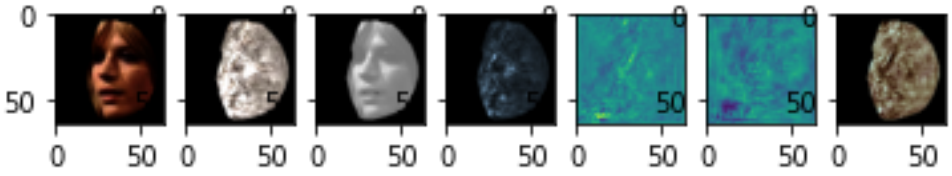

Figure 2b: Face decomposition maps that resemble BioFaceNet's Figure 1 after replacing shading with the grayscale image. From left to right the plotted maps are original RGB, reconstructed, RGB, shading, specularities, melanin, haemoglobin, and diffuse albedo (as in Figure 1 of the original BioFaceNet paper).

Reproducing the authors' older paper

To reproduce the results in Table 1, we need to reproduce the authors' older paper, Decomposing Multispectral Face Images Into Diffuse and Specular Shading and Biophysical Parameters (Smith 2019, available at https://arxiv.org/pdf/1902.06557.pdf) since BioFaceNet's Table 1 result is the error difference between the results in this older paper versus BioFaceNet. Now, the authors use the ISET database instead of CelebA since ISET contains spectrometer measurements of faces. ISET does not contain any real RGB images.

In (Smith 2019), they make the same physics model assumptions about haemoglobin, melanin, specularity, and diffuse masks as BioFaceNet but instead of using a deep learning network to learn them, they use least squares minimization to estimate them directly.

It is hard to write the least squares minimization (see Eq. 3 in (Smith 2019)) in closed form. Assuming the problem is convex, one solution is to minimize iteratively with SGD until convergence and then export these optimal torch weights, and that's the implementation used here.

This method produced the image below for the first ISET sample.

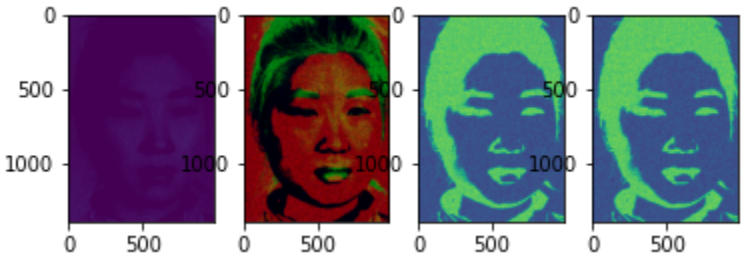

Figure 3: Face decomposition based on least squares minimization. From left to right the plotted maps are diffuse, specularities, melanin, and haemoglobin.

Although the results appear far from the actual diffuse, specular, melanin, and haemoglobin maps, this optimization procedure appears to have recovered some meaningful output. This also could mean there is no major error in the PyTorch implementation of several face reconstruction methods which are also used both in the Figure 3 above and in the BioFaceNet reimplementation in Figure 2 above.

If the BioFaceNet model worked, we would have D65-lighted the spectrums to produce an RGB image and passed that into BioFaceNet, comparing the maps below with those obtained from BioFaceNet. Taking the RMSE of the image differences would reproduce the second row of Table 1 in the paper.

Conclusion

The actual subset of the data the authors used is unclear, as they read from a matfile which is not present in the GitHub repo. When describing their CelebA data in the BioFaceNet paper, the authors cite a paper whose augmented CelebA dataset is on Google Drive (including parameters needed to compute shading), which is the one used

in this reimplementation. It wouldn't be possible to use the official CelebA data since there are no shading/harmonics parameters in it. If the authors obtained their shading data differently, this may have a large impact on why the results are not the same. Moreover, there are some potential overfitting issues about the pseudo-ground truth data mentioned in the first section here.

The authors' older paper based on least squares minimization appears promising, and further work on it could shed light on why BioFaceNet doesn't converge.